

DOCUMENT RESUME

ED 276 763

TM 860 714

AUTHOR McLean, James I.
 TITLE Formative Evaluation of the Effectiveness of Texts.
 PUB DATE [86]
 NOTE 33p.
 PUB TYPE Reports - Research/Technical (143)
 EDRS PRICE MF01/PC02 Plus Postage.
 DESCRIPTORS *Comprehension; *Formative Evaluation; Higher Education; *Instructional Materials; Judges; *Material Development; Multiple Choice Tests; *Predictive Validity; Pretests Posttests; Rating Scales; Readability Formulas; Regression (Statistics); Student Characteristics
 IDENTIFIERS *Written Instruction Stylistic Evaluation

ABSTRACT

The first purpose of this study was to examine the relationships between changes made during the development of written instruction and the effectiveness of the materials when presented to students. For this purpose, the Written Instruction Stylistic Evaluation was used to assess the comprehensibility of passages before administration of the passages to students. The second purpose of this study was to assess the degree to which the rewriting efforts of experienced writers are successful in improving the comprehensibility of written instruction. The instrument for this study was formulated by translating modern concepts of reading comprehension and instructional development into 12 guidelines for writers in the form of anchored rating scales. Regression equations predicted much of the variance in the student scores. Thus, this study found strong predictive relationships between the changes made during instructional development and the students' success in using the resulting materials. However, the optimum regression equation was different for each outcome measure. (JAZ)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED296763

Formative Evaluation

1

Formative Evaluation of the
Effectiveness of Texts

James I. McLean

Academy of Health Sciences, U.S. Army

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

J.I. McLean

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

Running head: FORMATIVE EVALUATION OF THE EFFECTIVENESS OF TEXTS

The views expressed here are my own and do not purport to reflect the position of the Department of the Army or the Department of Defense. This paper is based in part upon research for my dissertation at the University of Texas in Austin. The chairs were Carol E. Kasworm and Diane L. Schallert.

BEST COPY AVAILABLE

TM 860 714

Abstract

This study examined the ability of writers and judges to assess the effectiveness of texts when rating scales were used. Writers and judges were asked to assess factors affecting the comprehensibility of passages before and after revision. The predictive validity of these assessments was then tested by administering the materials to students.

The instrument for this study was formulated by translating modern concepts of reading comprehension and instructional development into 12 guidelines for writers in the form of anchored rating scales. Regression equations predicted much of the variance in the student scores. Thus, this study found strong predictive relationships between the changes made during instructional development and the students' success in using the resulting materials. However, the optimum regression equation was different for each outcome measure.

James I. McLean
8707 Derringer Lane
San Antonio, Texas 78217

Formative Evaluation of the
Effectiveness of Texts

There is much concrete evidence that texts are important in education. Consider libraries, copying machines, bookstores, text selection committees for school systems (and the advocates who hope to influence the text selection process), and the brochures produced by publishers to sell their wares. Your own formal educational experiences may also indicate that texts often prescribe the curriculum. On the basis of several studies, Komoski and Woodward (1985) have concluded that 90 to 95 percent of instructional time is spent using instructional materials including texts. Even in the context of self-directed learning, books are rated as "extremely important" resources by 71.2 percent of active learners (Penland, 1977).

Since texts are so important, it is understandable that writers of instructional materials would continually seek strategies and guidelines which might help to improve and assure the comprehensibility and learning effectiveness of materials under development. Fortunately, there is a growing context of research and theory to assist instructional developers as they consider strategies and guidelines for writing. This context is derived from improved theories of reading comprehension, better models of

cognition, highly explanatory models of writing processes, and experience in the actual testing of instruction with students. This context should be helpful in identifying linkages between the changes made by writers and editors and the potential success of students in using resulting materials. However, these relationships are still poorly understood.

Jenkins' Tetrahedron

Figure 1 is adapted from Jenkins (1979) and Bransford (1979) to reveal variables in instructional success. The purpose of presenting it here is to provide a framework to consider guidelines for improving instructional materials. This figure is a tetrahedron, a pyramid with a triangular base and triangular sides. Its four corners are (1) the characteristics of the learners, (2) the demands placed upon the learner and the activities which the student performs in efforts to learn, (3) the materials used in the learning process, and (4) the outcomes produced by the learning efforts and the criteria by which the outcomes are measured.

Insert Figure 1 about here

The first corner of the tetrahedron emphasizes the characteristics

of the learner. One very important way in which learners vary is in their knowledge. This includes both their general knowledge of the language and knowledge specifically related to the subject of a text. A second important characteristic of learners is their level of motivation to learn successfully. Research has shown that with sufficient motivation, students may overcome factors such as sentence complexity to learn well from texts. Other characteristics varying among learners may include aptitude and age-related variables such as vision. Some writers would also emphasize cultural traits of learners.

One characteristic of learners which I like to emphasize is their purpose in reading (Diehl and Mikulecky, 1981). For example, a student may want to merely assess a text to see if it will be useful later as a reference or as material to be learned later. Second, a student may be reading to do something, as a student may refer to a text to perform a medical laboratory procedure. Third, a student may be reading the text in order to learn.

The second corner of the tetrahedron features demands placed upon students and the activities used by students to learn successfully. For example, an instructor may require that a student read a text only for familiarity or read it to apply principles. A student may skim a text, memorize extensive portions, or read deeply

to construct a lasting personal schema. I prefer to view the second corner of the tetrahedron in relationship to the student's purpose in reading. Thus, if a student is reading a text to assess, do, or learn, her activities may be somewhat different for each purpose. To assess a text, the student may read the title and the table of contents or check the index to see if a certain topic is included. To do something, the student may open the text to a specific procedure and read each step before performing it. In this instance, the text becomes a job performance aid. To learn from a text, the student will probably read more carefully and deeply. He may start at the front and read cover to cover or read proactively according to his own plan of study.

The third corner of the tetrahedron consists of the text or other materials used in learning. This corner is the chief emphasis of this paper, but it may be misleading or futile to consider any corner in isolation from the others. For example, the design of the text should consider the purposes and activities of the student. So that a text may be easily assessed, it should include a good title, an overview, frequent summaries, a table of contents, and an index. If a text is to be helpful in performing procedures, it should closely correspond to the job task and include any needed algorithms, graphics, tables for quick retrieval of data, and an

index. If the text is for learning, a number of different stylistic devices may be needed, such as adjunct questions, objectives, and organizers. Of course, texts must often be suitable for all three purposes--assessing, learning, and doing (Diehl and Mikulecky, 1981).

The fourth corner of the tetrahedron--after learner characteristics, demands and activities for the learner, and the nature of the materials--is concerned with the outcomes and the criteria used to assess the outcomes. The outcomes may include comprehension or understanding of the materials with or without a requirement for later memory. The outcomes may also include memory, both short term and long term. In some cases the student will be able to transfer learning to solve higher-order problems. As appropriate, the student may be able to perform the steps of a procedure.

Formative Evaluation at the Mountaintop

My primary interest is in the third corner of the tetrahedron, the nature of the materials used in instruction. Specifically, I am interested in the evaluation of materials as they are developed.

One question considered in formative evaluation is, "Do my materials teach?" In other words, can the students learn by using my materials?

In some cases, this may be a very important question. If a

group of students have previously been unable to apply an important concept and if new materials bridge the gap, the materials are valuable. A typical way of answering this question of effectiveness is to find a group of nonmasters (people who fail a criterion pretest), expose them to instruction, and see if a certain percentage, maybe 80 percent, can pass a similar or identical criterion posttest. In other words, find students unfamiliar with the content, instruct them, and determine that they then know some of the content. I call this the mountaintop approach to formative evaluation because it measures achievement of objectives that were previously unmet.

Campbell and Stanley (1963) refer to the typical method used in the mountaintop approach as the one-group pretest-posttest design. In many cases, there are rival sources of difference which can explain the improved performance of the students. These can include the effects of pretesting, regression toward the mean, and uncontrolled events that occur between the pretest and the posttest, such as assistance from other students or the experimenter. Also, it is risky to generalize the results. In the natural field situation, for example, there may be different students and no experimenter to motivate the students.

An additional problem is that there may be many instructional objectives for which any coherent text will be adequate. If the

instructional goal is to learn Subject X, consisting of straightforward verbal information and simple concepts, it is probably redundant to perform formal trials. Given any reasonably coherent text about Subject X, enough time to study it, and enough time to take a test, a group of nonmasters will be able to pass the test. Formal trials in this case would be an expensive waste.

Another problem is that there is a tendency to emphasize those instructional objectives and test items which are most likely to result in "successful" validation. That is, there is a tendency to emphasize material which is unfamiliar but easily learned, including simple rules and facts. There are many important learning goals that can not be so quickly achieved in a short period of instruction.

Formative Evaluation on the Mountainside

Before we evaluate instruction to see if we are at the top of the mountain, we must also learn whether we're climbing properly. That is, are we using appropriate strategies to get to the top of the mountain, to improve the effectiveness of instruction? For such reasons, it is often preferable to emphasize a different question in formative evaluation, "How can I make my materials teach better?" To answer this question, the mountaintop approach must be supplemented or preceded by a mountainside approach, that is, a diagnostic approach. Diagnostic approaches to formative evaluation

typically involve editorial review, content review, and actual testing of materials with students. The emphasis in testing, however, should be on qualitative information which might be specifically useful in revision and not on the absolute amount of learning in the "mountain." Most instructional developers spend far more time and effort on the mountainside than they do on the mountaintop.

Both the power and the limits of editorial review have been brought to light in a study by Flower, Hayes, Carey, Shriver, and Stratman (1986). When expert writers revised, they were able to fix or remove 91 percent of problems placed by the experimenters. Novices were able to remove only 64 percent of the problems. Due to the limitations of any one reviewer, it is probably wise to include at least two persons other than the writer in the review process.

There may be learning problems produced by certain texts which can not be identified by editorial review. As Jenkins' tetrahedron shows, there are very complex relationships among the different variables of instruction and the outcomes produced. For such reasons, it is often useful to test materials upon students who are representative of the target population, including both masters and nonmasters. For diagnostic purposes, there is little need to distinguish nonmasters from masters or to compare pretest and posttest performance.

An intriguing alternative to established methods of testing instructional materials was developed by Swaney, Janik, Bond, and Hayes (1981). They successfully improved the comprehensibility of an insurance policy by taking reading protocols with a few subjects and then providing the results of the protocols to experienced document designers. In some respects, this newer approach resembles one-to-one testing, the initial phase of formative evaluation as described by Dick and Carey (1985). With single individuals, it may be possible to use a protocol method to track the student's line of thought while receiving short instructional segments.

When larger groups are tested with longer instruction, questionnaires or interviews should be used to find sources of difficulty. Students can make helpful suggestions for revision, such as the need to define a term better or identification of test items needing revision. Tabulation of answers to test items can be a valuable aid in detecting instructional weaknesses.

I previously mentioned some scientific limitations of the mountaintop method of testing instruction on nonmasters. The diagnostic approach is even less qualified as an experimental design. More often than not, a diagnostic procedure should be classified as a one-shot case study. On the other hand, we are often more interested in diagnosis and treatment--the diagnostic

or mountainside approach which detects problems and suggests solutions--than we are in being certified as sick or healthy without identification of cause--the mountaintop approach.

Problems With Readability Formulas

Some developers of instructional materials have used readability formulas as a device for adapting materials for a given target audience. Recently, the International Reading Association and the National Council of Teachers of English issued a position statement regarding the use of readability formulas by publishers of instructional materials. According to this statement, a major problem with readability formulas is that they usually consider only word and sentence length. However, there are many other factors which can affect the readability or effectiveness of materials. These factors include "syntactic complexity of sentences, density of concepts, abstractness of ideas, text organization, coherence and sequence of ideas, page format, length of type line, length of paragraphs, intricacy of punctuation, and the use of illustrations and color" (International Reading Association, 1984/1985).

Another major problem with the use of readability formulas is that they encourage belief in a fallacy, as illustrated in Figure 2. The first part of this fallacy is that the reading ability of a student can be accurately categorized by a single number, the student's grade level.

Insert Figure 2 about here

This is an oversimplification which might be useful for some limited purposes. The second part of this fallacy is that the difficulty of a text for students in general can also be measured with a single number, the reading grade level as computed by formula. In conjunction, the two parts of this fallacy encourage us to believe that the average word length and sentence of the text should be adjusted to yield a formula score corresponding to the grade level of the student. Although a student's reading ability has many elements or facets and the difficulty a student encounters in reading a text involves many variables, readability formulas may lead us into a trap in which we consider only two characteristics of texts. The limitations of this approach are clear if we refer back to Jenkins' tetrahedron and see how much is left unconsidered.

A number of studies have demonstrated that when materials are revised to shorten sentence length and to substitute familiar words for unfamiliar words, the gains in comprehension are often insignificant or unimportant. Reviewing 36 such experimental studies, Klare (1976) concluded that improvements in comprehension were achieved in 19 studies,

no significant differences were found in 11 studies, and the remaining 6 studies produced mixed significant and insignificant results. Three more recent studies have failed to support revision according to readability formulas (Duffy and Kabance, 1982; Johnson and Otto, 1982; Walmsley, Scott, and Lehrer, 1981).

Jenkins' tetrahedron has helped us to consider the many variables involved in comprehension and memory. It is no surprise then that readability formulas have not become a magic formula for making our materials more effective. After all, readability formulas consider only two language variables in texts. In my own research (McLean, 1985), I considered a wide range of stylistic variables that might affect the comprehensibility of prose.

Method

The first purpose of the study was to examine the relationships between changes made during the development of written instruction and the effectiveness of the materials when presented to students. For this purpose, writers and reviewers used a new instrument, the Written Instruction Stylistic Evaluation (WISE) scales, to assess the comprehensibility of passages before administration of the passages to students. The second purpose of this study was to assess the degree to which the rewriting efforts of experienced

writers are successful in improving the comprehensibility of written instruction.

Subjects

This study involved three categories of subjects:

Writers. I wrote letters to 799 members of the Association for Educational Communications and Technology and the National Society for Performance and Instruction and contacted 55 other colleagues in instructional development to ask them to serve as writer-subjects. Those volunteering (150) were randomly assigned one of two passages to revise. However, only about one-half of those volunteering actually completed the task. This produced 42 usable revisions of the first passage and 31 usable revisions of the second passage.

Judges. Five experienced education specialists in the Course Development Division, Academy of Health Sciences, participated in the development of the passage rating instrument. Five more participated in a corollary study to examine the validity of the vocabulary scale. Five of these first 10 served as judges to use the Written Instruction Stylistic Evaluation (WISE) scales to rate both the original and the revised passages. Six education specialists served as content judges, each rating half of the passages according to whether each passage provided sufficient information to answer each item on the corresponding test.

Students. Three incoming classes of military trainees at the Academy of Health Sciences participated. Passages were randomly assigned with the goal of administering the two original passages to 60 students each and the revised passages to 12 students each. Memory and comprehension tests from 962 of these subjects were usable for analysis. Aptitude data were successfully retrieved for 865.

Materials

Original passages. For this study, I sought two original passages which were (1) relevant in content for the early training of medical soldiers, (2) sufficiently difficult to create a comprehension gap which might be bridged by appropriate changes in the passages, (3) believable or "natural" as texts that medical soldiers might reasonably encounter during their training, (4) short enough to present a reasonable writing task for volunteer subjects, (5) long enough to represent text effects beyond the word and sentence level, (6) approximately equal in length to each other, and (7) complete, that is, able to stand on their own without appearing to be disjointed fragments of text. The two passages used were selected on the basis of my prior experience as a correspondence course writer for the Army Medical Department. One passage, about the role of the medic in combat, was an introduction to a manual for medical soldiers. The other passage, about red cell components for

transfusion, was an extract from a blood banking manual. This second passage was more difficult in terms of technical vocabulary and concepts than those typically encountered by entrance-level medical soldiers.

Written Instruction Stylistic Evaluation (WISE) scales. The WISE scales were designed to measure prose factors related to comprehensibility. Each scale, resembling a thermometer, was on a separate piece of paper. There were descriptive anchors attached to points of each scale from 10⁰ to 90⁰.

Each WISE scale was headed by a question:

1. How familiar is the vocabulary?
2. Are sentences simple or complex?
3. Are examples, metaphors, and analogies used well?
4. Is good use made of illustrations?
5. How well is material related to the reader's knowledge?
6. How much explanation is provided?
7. How much variety is there?
8. Are objectives or purposes indicated?
9. Is student practice provided?
10. Is the organization easy to see?
11. Are references clear?
12. Do you read your own writing? (This scale was used by writers only.)

Tests. One 24-item multiple-choice test was developed for each of the two original passages. Each test consisted of four levels of six items each. In concept, these levels were adapted from a text edited by Pennock (1979), which discussed instructional strategies for comprehension in terms of four linguistic levels--the word, the sentence, the paragraph, and beyond the paragraph. The levels in the tests were labeled as follows: "WORD MEANINGS," which tested knowledge of the meanings of words used and not generally defined in the original passages but which may have been defined in some revisions; "SENTENCE MEANINGS," which tested knowledge of relationships within single sentences in the original passages; "LOCAL MEANINGS," which tested relationships of six pairs of contiguous sentences located in each of the original passages; and "GLOBAL MEANINGS," which tested judgment of the author's purpose, the best title, the best summary, the worst analogy, and the best structural depiction of the macrostructure of the text. In general, the test items avoided using the exact wording in the original passages, but some substantive words were repeated in some test items. The danger of testing verbatim processing rather than comprehension was present mainly with items testing SENTENCE MEANINGS. There was a clear inferential component to the other levels of the test.

Since a copy of the appropriate test was provided to each of the revising writers, there was an opportunity for each writer to "teach the test." This type of test is not typically encountered by instructional developers or students. The tests were specifically designed for the purpose of this experiment. A more important criterion than level of meaning for ordinary tests would be the explicit relevance of the test item to knowledge needed for job performance. Multiple-choice tests are commonly used in military training, but they are more frequently sequenced in relationship to tasks for job performance.

Procedure

This study consisted of five phases:

1. Development of the WISE scales. These were anchored rating scales, each with a graphic of a thermometer to convey the concept of equal-interval measurement. Each scale was developed to reflect an aspect or cluster of concepts, consistent with modern views of comprehensibility and instructional development. By the use of this instrument, changes made by writers could be classified and sufficiently quantified for analysis.
2. Assessment of a passage by each writer. A large number of writers were included in order to achieve a naturalistic sampling of writing styles.

3. Rating of original and revised passages by judges. All original and revised passages were rated on the first 11 WISE scales by each of five judges. The adequacy of the content of each passage was later rated by three out of six judges.

4. Administration of the passages to students. In addition to considering a wide range of stylistic variables, my study differed from other readability studies in one other major respect. It employed a range of outcome measures and a measure of student aptitude. Each test of student performance had four different levels, and each test was administered in both an open-book condition and a closed-book condition.

5. Analysis of the data. A variety of statistical analyses were used to examine the reliability and validity of the WISE scales. By validity, it is meant that the ratings of the writers or the judges would actually have some predictive value for determining the effectiveness of written instruction. To the extent that the scales were valid, this study also examined the relationships between changes made by writers and the effectiveness of the materials with students.

In a sense, each of the WISE scales represented a hypothesis for testing. For example, as a writer made changes in the vocabulary of a passage and registered the changes by marking the "thermometer"

for the vocabulary scale, the implied hypothesis was that the writers' ratings would be predictive of changes in the effectiveness of the materials. Likewise, considering the mean of the five ratings of the judges on an identical WISE scale, the implied hypothesis was that the mean ratings would also be predictive of effectiveness with students.

In addition to the group of hypotheses implied by the WISE scales, there was one other major hypothesis considered in this study. This is related to the question of whether professional writers as a group can improve either of two different passages. The null hypothesis was that there would be no difference between the revised and the original passages in terms of their effectiveness with students.

Results

A different regression equation was optimum to predict each dependent measure. Previously, Bormuth (1969) found that readability of materials could best be predicted with equations consisting of more than two independent variables. My study demonstrated that the predictive equations may also vary, depending upon choice of outcome measure. This indicates that the usefulness of a specific revision strategy may vary depending upon how the text is to be used by the readers. No revision strategy can be expected

to improve every outcome measure with every student. Nonetheless, the following revision strategies are often effective:

(1) Verify content. Content ratings by three judges were the most consistent predictor of the effectiveness of the passages.

(2) Simplify vocabulary. Although this variable appeared in only 1 of 10 regression equations, it is probably still a useful strategy. Writers should remember that students are expected to learn the vocabulary appropriate for a subject area. They can not do this if difficult words are completely avoided.

(3) Use simple sentences. Some partial correlations indicated a positive role for shorter sentences. Others indicated a negative role. This item did appear in a regression equation to predict total memory scores. You should consider this strategy for improving the effectiveness of written instruction. However, you also run a risk of making some relationships unclear, especially if you delete such words as since and because.

(4) Use examples, metaphors, and analogies. At times, examples, metaphors, and analogies can improve memory of concepts, although they may also slow comprehension. Other devices--such as pretests and special introductions--should also be considered to tap

(5) Specifically state objectives. Specifically stated objectives may be useful except when memory for global meanings is important. Students tend to do as they are told. If you tell them to learn specifics only, they will tend to do just that.

(6) Improve the clarity of references. Five of 10 regression equations showed important and positive roles for steps made by writers to improve the clarity of relationships between one part of a text and other parts.

The following strategies produced mixed effects:

(1) Use illustrations? Graphic illustrations do not always increase the effectiveness of instructional materials. They should be used when there is a clear need for information which can not be conveyed easily in words.

(2) Use appropriate elaboration? As a general strategy, seeking just the right level of elaboration tends to enhance comprehension of word meanings but may interfere with other levels of comprehension and memory.

(3) Provide variety and informality? Increasing the variety or informality of a passage is not a dependable way of making it more memorable or comprehensible.

(4) Improve the clarity of organization? There are mixed results when efforts are made to improve the clarity of organization. Changes in clarity of organization are almost

necessarily linked to changes in the overall meaning of the text. By implication, there are also changes in instructional objectives and a need for changes in the outcome measures. Since the design of my study did not allow for changes in the outcome measures, there was a penalty for changes in the meaning of the instruction.

The scale assessing the level of student practice was insignificantly or negatively related to student outcome measures. This was probably due to the very short times allowed for students to read the passages.

The writers felt they had radically improved both of the original passages with respect to virtually all of the WISE scales. The judges agreed that the passages were improved but more modestly than thought by the writers. Overall, the content judges felt that the revised passages were not quite as complete as the original passages. In terms of student performance, the revised passages were more effective than the original in conveying word meanings. Considering total test scores, however, there was no significant difference between the revised and the original passages.

Conclusions and Implications

There is one additional way to improve the quality of our instructional materials which I have not yet mentioned. This is by improving our writing skills. We can do this by improving our

knowledge of our students, improving our knowledge of communications, and improving our knowledge of our subject matter. A good system for instructional development will ensure that we get the appropriate feedback from students and reviewers to help us improve in these areas. These types of knowledge will help us to select the best words, analogies, metaphors, examples, and graphics to reach our students. As we become more skilled in formative evaluation, we will be better able to detect and diagnose problems in instruction. As we improve our other writing skills, we will be better at devising effective solutions to the problems detected in the diagnostic approach suggested here.

I think future research in instructional development will include two major strains--quantitative studies of large cross-sections of Jenkins' tetrahedron (such as this study) and qualitative approaches emphasizing the protocol method to examine writing and learning processes. Properly done, these methods will complement each other. The qualitative studies are of special value in understanding the results of the quantitative ones.

References

- Bormuth, J. R. (1969). Development of readability analyses. The University of Chicago. (ERIC Document Reproduction Service No. ED 029 166)
- Bransford, J. D. (1979). Human cognition: Learning, understanding and remembering. Belmont, CA: Wadsworth.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Dick, W., and Carey, L. (1985). The systematic design of instruction (2nd ed.). Glenview, IL: Scott, Foresman.
- Diehl, W., & Mikulecky, L. (1981). Making written information fit workers' purposes. IEEE Transactions on Professional Communication, PC-24(1), 5-9.
- Duffy, T.M., and Kabanec, P. (1982). Testing a readable writing approach to text revision. Journal of Educational Psychology, 74, 733-748.
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and strategies of revision. College Composition and Communication, 37, 16-55.
- International Reading Association. (1984/1985). IRA, NCTE take stand on readability formula. Reading Today, 2(3), 1.

- Jenkins, J. J. (1979). Four points to remember: A tetrahedron model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 429-446). Hillsdale, N. J.: Lawrence Erlbaum.
- Johnson, L. L., and Otto W. (1982). Effect of alterations in prose style on the readability of college texts. Journal of Educational Research, 75, 222-229.
- Klare, G. R. (1976). A second look at the validity of readability formulas. Journal of Reading Behavior, 8, 129-152.
- Komoski, P. K., & Woodward, A. (1985). In Jonassen, D. H. (Ed.), The technology of text: Principles for structuring, designing, and displaying text (pp. 396-417). Englewood Cliffs, NJ: Educational Technology.
- McLean, J. I. (1985). Formative evaluation of the comprehensibility of written instruction. Dissertation Abstracts International, 47, issue 2. (University Microfilms No. 86-09,547)
- Penland, P. (1977). Individual self-planned learning in America. Washington, DC: U.S. Department of Health, Education, and Welfare.
- Pennock, C. (Ed.). (1979). Reading comprehension at four linguistic levels. Newark, DE: International Reading Association.

Swaney, J. H., Janik, C. J., Bond, S. J., & Hayes, J. R. (1981).

Editing for comprehension: Improving the process through reading protocols. Washington, DC: American Institutes for Research.

Walmsley, S. A., Scott, K. M., & Lehrer, R. (1981). Effects of document simplification on the reading comprehension of the elderly. Journal of Reading Behavior, 13, 237-246.

Figure Caption

Figure 1. A tetrahedron used to illustrate the variables involved in comprehension and memory.

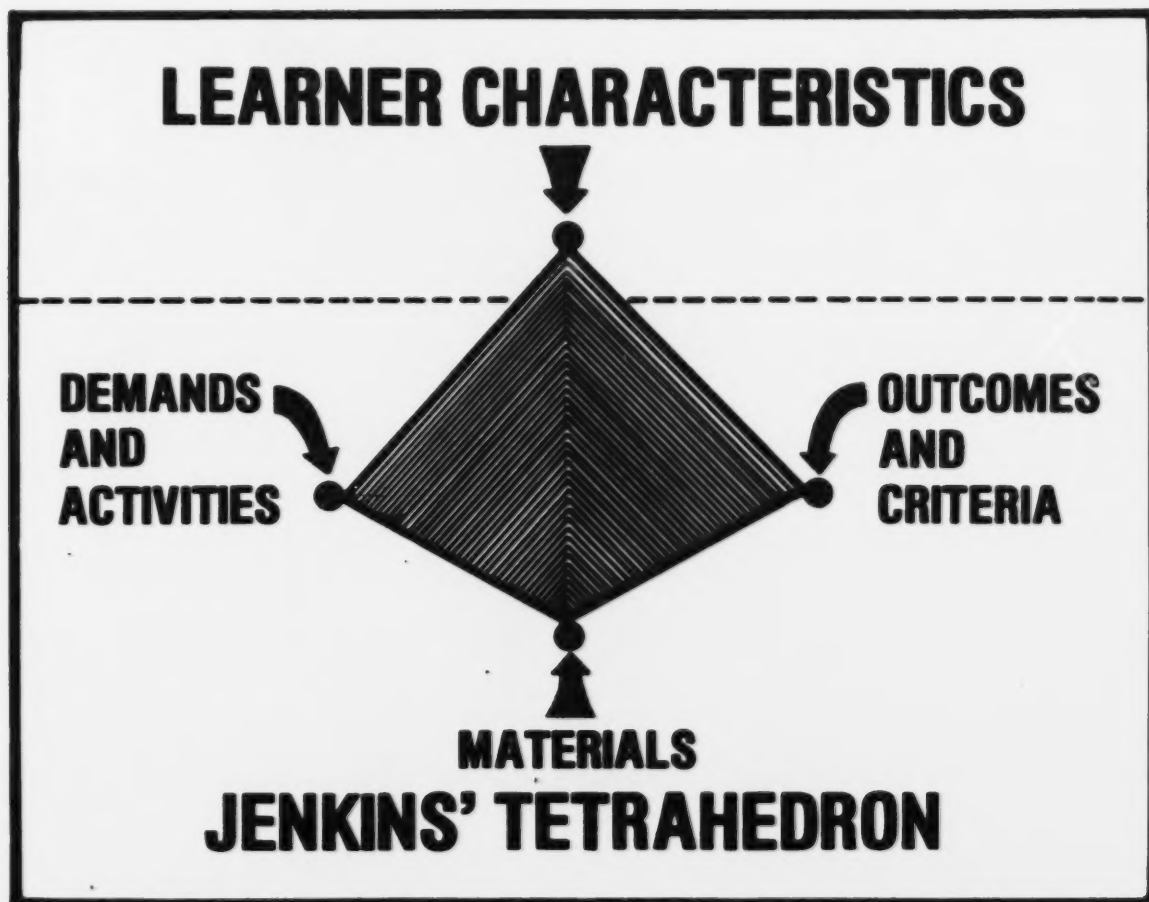


Figure Caption

Figure 2. Fallacy consisting of two parts: (a) assessing reading ability with a single number and (b) assessing text difficulty with a single number.

